5.5 A 64B CPU Pair: Dual- and Single-Processor Chips

E. B. Cohen[1], N. J. Rohrer[1], P. Sandon[1], M. Canada[1], C. Lichtenau[2], M. Ringler[1], P. Kartschoke[1], R. Floyd[1], J. Heaslip[1], M. Ross[1], T. Pflueger[2], R. Hilgendorf[3], P. McCormick[1], G. Salem[1], J. Connor[1], S. Geissler[1], D. Thygesen[1]

[1]IBM, Essex Junction, VT
[2]IBM, Boeblingen, Germany
[3]IBM, Austin, TX

A chip with two 64b PowerPC™ microprocessors, each with a 1MB dedicated L2 cache and a single shared high-speed processor-interconnect (PI) [1] bus is created. A second single-processor chip with a 1MB L2 cache is also created with a different performance/power optimization. The chips are built in 90nm dual strained-silicon SOI technology [2] using 10 layers of copper interconnect and low-k dielectric.

The PPC970MP dual-processor chip (MP) in Fig. 5.5.7 consists of 2 processor units (PUs) that are mirrored, and a common region. The I/O and PLL circuitry reside in the common area. The design of the PU core is an extension to a previous PowerPC™ design [3]. This core is the basis for both the MP and the PPC970GX single-processor chip (SP). The PU contains a 64kB L1 instruction cache, a 32kB L1 data cache supported by 2 load-store units, and a unified 1MB L2 cache. Each PU can dispatch up to 5 instructions per cycle and issue one instruction per cycle to each of its execution units, of which there are 2 integer, 2 floating point, 2 load/store, 2 single-instruction, multiple-data execution units and 2 additional units that execute control operations.

Both chips have a single PI [1] off-chip bus, consisting of 2 unidirectional source-synchronous single-ended links. The 36 data bits of each link are encoded into 44b for parity and to minimize simultaneous switching noise. The circuit in Fig. 5.5.1 generates an adjustable reference voltage (vref) by shorting the terminated differential inputs, clkin and clkin_b, together through series resistors and passgates. Low-pass filtering allows vref to ride common-mode noise, improving the eye.

Vref windage adjusts vref up or down to compensate for asymmetrical voltages or input waveforms. Tunable termination windage bits select which 14 terminating resistors are enabled, maintaining constant input resistance.

The bit rate per channel scales with core frequency in a 1:2 ratio, as it does in the Power5™ design [1], to keep pace with data demand. At 3GHz core frequency, the PI provides 10.7GB/s data bandwidth as well as address and control overhead.

The MP is divided into clock and voltage domains, as shown in Fig. 5.5.2. Processor 0 (P0), the I/O control and the single PLL share the Vdd0 supply and clock mesh 0. A separate mesh, mesh 2, covers the L-shaped region where the PI receivers are placed. Mesh 2 is also powered by Vdd0. P1 is supplied by Vdd1 and clocked by mesh 1.

The arbiter, shown in Fig. 5.5.3, controls the shared off-chip bus of the MP. The chip sends a data stream of 2 to 34 beats, 36 logical-bits wide. The first 2 beats are a header defining the packet function. The arbiter defines the path between the individual processor and the PI bus for the packet, e.g., if P0 wins arbitration, its packet flows through latch Lt1 only. The packet from the losing processor, P1, has its header bits stored in Lt1 and Lt0 latches of P1. The arbiter signals P1, halting its data transfer to the bus. In round-robin fashion, P1 wins the next arbitration. P1 shifts out the contents of its Lt0 and Lt1 while a signal is sent to P1, initiating transfer of the remainder of its packet.

A backoff method handles bus-contention issues. The bus pacing parameter (BPP) of each processor defines the number of internal bus clocks between commands to the bus. The BPPs can be configured to change independently, based on the number of retries each bus receives.

Figure 5.5.3 shows that arbiter transactions across voltage-domain boundaries include fencing circuitry described below. Arbiter transactions across clock-mesh boundaries include timing allowances.

A level-shifting and logic fencing network, Fig. 5.5.4, is placed in series with each signal that crosses between power domains. The first inverter in each receiver is ratioed to lower the threshold by about 100mv in order to provide a solid level to the receiving domain and reduce cross-over current during voltage transients. A negative active 'fence_b' signal is provided to ensure that all inputs receive a known state as one PU is faded offline.

To account for voltage and clock skew between domains, ±98ps is added to the timing of paths that cross between clock mesh 0 and mesh 1. ±40ps is added to paths that cross between clock mesh 0 and 2. Inter-mesh clock skew and local-voltage variation are measurable through the use of 4 "skitter" [4] macros per chip. Each voltage rail is monitored on-chip by dedicated Kelvin probes. The Kelvin monitors of the SP measure voltage at the silicon at 32 sites across the chip. Figure 5.5.5 shows the connection of the probe to output pins through selectable pass-gate circuitry.

When the MP initiates power-on, signals running between power planes are gated to allow independent PU initialization. P0 and the front side bus are initialized first. P0 then starts execution. Vdd1 is powered up next and P1 is initialized. The fences between power planes then allow signals to pass and P1 begins execution. The chip supports dynamically stopping P1, depowering it, and restarting it while P0 is running.

Both chips utilize multiple strategies to reduce power consumption during intervals of off-peak computing demand, including the power-tuning [5] facility that allows the core frequency to switch between full, 50%, and 25% of maximum frequency in a few cycles. Powertuning has been extended to allow toggling between full and 25% speed in a single tuning sequence. Voltage can be scaled during any of the powertuning modes. While both PUs are operating, the 2 cores have identical frequencies at all times, including the Deep-Nap mode, where the cores operate at 1/64 of their maximum frequency. Two other reduced clock-buffer activity modes, Nap and Doze, are also supported on both chips. The SP supports Very Deep Nap, which allows further voltage scaling and locks the state until release is received from the north bridge.

The operating range of the MP is 1.2 to 3GHz. Its estimated typical power consumption is 32W at 1.7GHz and 100W at 2.5GHz. The SP uses 16W at 1.6GHz and 85W at 3GHz.

Both chips are built in a partially-depleted 90nm SOI technology [2]. The NFET and PFET use a nitride layer to induce tensile and compressive strain on the underlying silicon to improve their respective mobilities. Eight of the ten layers of copper interconnect have a low-k dielectric. The SRAM cell size is 1.06μm^2. Additional technology and processor features are shown in Fig. 5.5.6.

References:

[1] D. Dreps, et al., "IBM Power5™ Bus Designs for On- and Off-Module Connections", Electrical Performance of Electronic Packaging, pp 173-176, 2004.

[2] H. Yang, et al., "Dual Stress Liner for High Performance sub-45nm Gate Length SOI CMOS Manufacturing," *IEDM Tech. Digest*, pp. 1075-1077, Dec., 2002.

[3] N. Rohrer, et al., "PowerPC 970 in 130nm and 90nm Technologies," *ISSCC Dig. Tech. Papers,* pp. 68-69, Feb., 2004.

[4] P. Restle, et al., "Timing Uncertainty Measurements on the Power5™ Microprocessor," *ISSCC Dig. Tech. Papers*, pp. 354-355, Feb., 2004.

[5] C. Lichtenau, et al., "PowerTune: Advanced Frequency and Power Scaling on 64-bit PowerPC Microprocessor," *ISSCC Dig. Tech. Papers*, pp. 356-357, Feb., 2004.

 1-4244-0079-1/06/$20.00 ©2006 IEEE

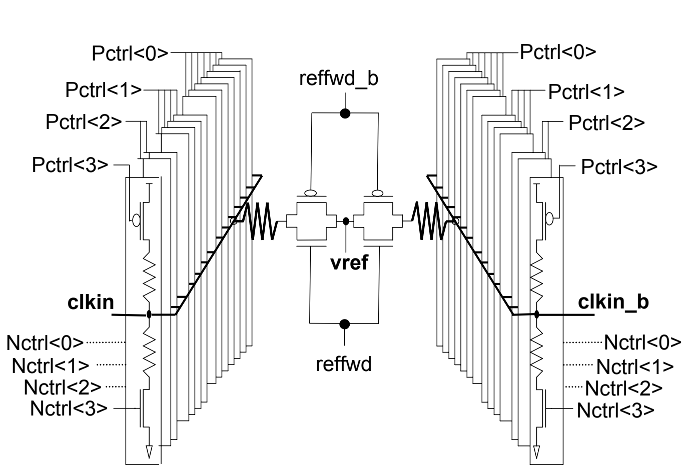

Figure 5.5.1: Vref forwarding circuit.

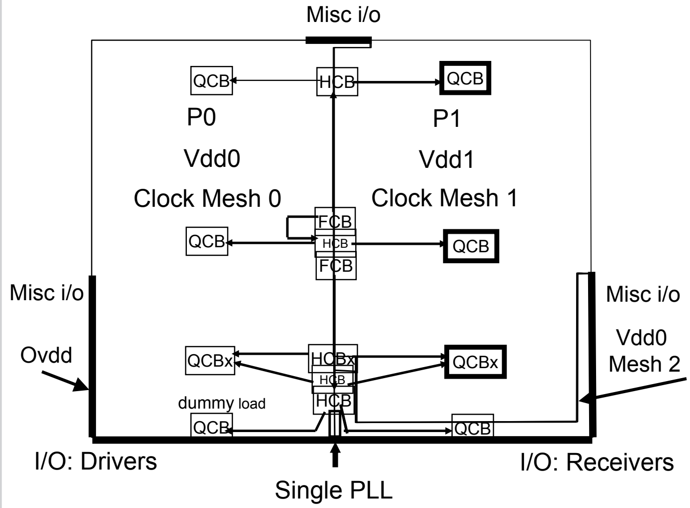

Figure 5.5.2: MP clocking and voltage domains.

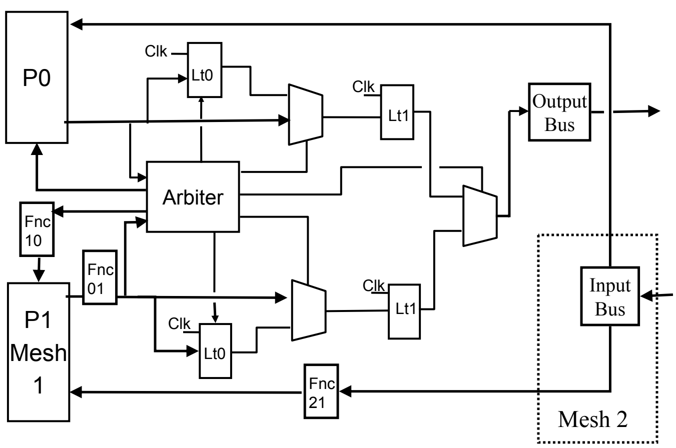

Figure 5.5.3: Arbitration block diagram.

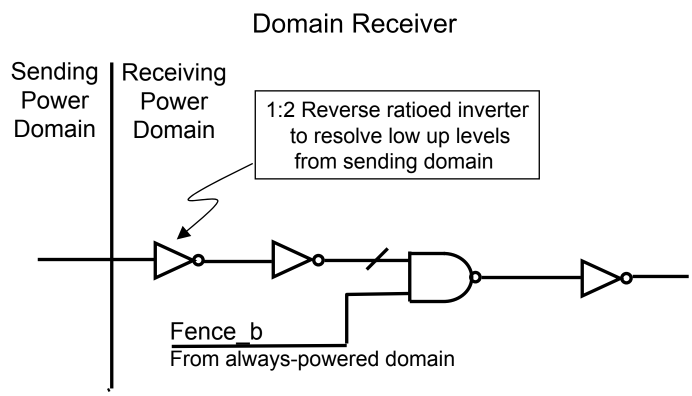

Figure 5.5.4: Domain boundary circuit.

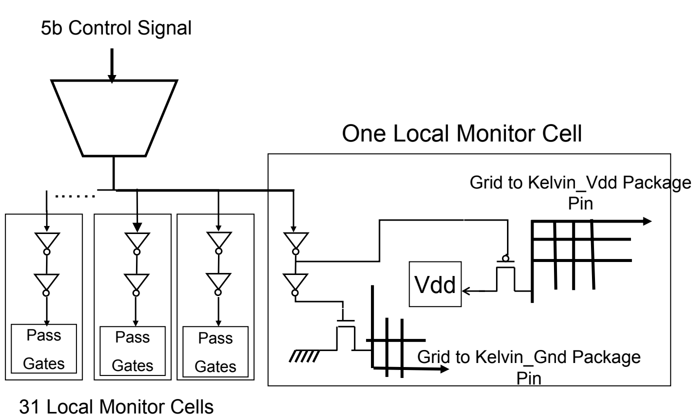

Figure 5.5.5: SP analog voltage monitor network.

	PPC970GX	PPC970MP
Technology	90nm CMOS w/ dual strained SOI, copper interconnects and low-k dielectric	
Gate Lpoly	46nm	
Tox	11.2, 15 & 22 Angstrom	
NFET/PFET I_{dsat}	900/490µA/µm @ 1.0V	
Metal Levels	10 (5-1x, 3-2x, 2-6x) – Low-k/FTEOS	
Die Image Size	79mm2	146mm^2
FETs	92.3 Million	183 Million
L1 Caches	64kB per core, instruction cache, w/parity 32kB per core, data cache, w/parity	
L2 Cache	1MB, unified cache w/ECC	1MB unified cache per core, w/ECC
Voltages	0.9V-1.4V core voltage 1.2-1.5V I/O	
Estimated Performance	1677 SPECint2000 @ 3GHz 19.5 SPECint_rate 2368 SPECfp2000 27.5 SPECfp_rate	1677 SPECint2000 @ 3 GHz 37.7 SPECint_rate 2368 SPECfp2000 48.8 SPECfp_rate
Package	25x25mm CBGA 575 pins on 1mm pitch (171 signals)	25x25mm CBGA 575 pins on 1mm pitch (181 signals)

Figure 5.5.6: Chip and technology features.

Continued on Page 641

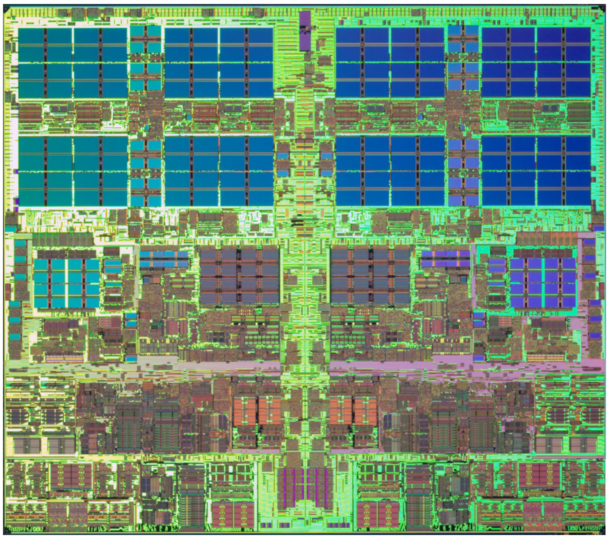

Figure 5.5.7: PPC970MP chip micrograph.